

Extracting Lexical Units for Identifying Specialized Semantic Frames¹

ANA OSTROŠKI ANIĆ

Institute of Croatian Language and Linguistics

MAJA LONČAR

Institute of Croatian Language and Linguistics

MARTINA PAVIĆ

Institute of Croatian Language and Linguistics

ABSTRACT

The paper discusses the idea of using parallel corpora for the extraction of terminological units used to identify semantic frames of aviation, as the first step in developing a method for determining specialized semantic frames. The process of compiling the parallel corpus of texts taken from the Eur-Lex database using Sketch Engine tools is described first. The methodology of extracting term candidates and their manual verification is then presented, as well as the process of extracting and verifying relevant verbs. The results of term verification are discussed, with the focus on the role of verbs and verbal collocations in identifying relevant semantic frames and frame elements.

KEYWORDS: aviation terminology, English-Croatian corpus, semantic frames, specialized knowledge, terminology.

ANOTACIJA

Straipsnyje svarstoma, kaip lygiagratieji tekstynai gali būti naudojami terminijos vienetams, taikomiems aviacijos semantiniams freimams nustatyti, atrinkti. Tai pirmoji metodo specializuotiems semantiniams freimams nustatyti kūrimo pakopa. Pirmiausia aprašoma, kaip, remiantis *Eur-Lex* duomenų baze ir naudojantis *Sketch Engine* priemonėmis, rengiamas lygiagratusis tekstų rinkinys. Tada pateikiami terminų variantų išrinkimo ir jų patikrinimo rankiniu būdu metodai bei

¹ This paper is based on the talk “From parallel corpora to specialized semantic frames”, given at the 3rd international conference on terminology *Scientific, Administrative and Educational Dimensions of Terminology* in Vilnius, 17–18 October 2019. This work has been fully supported by the Croatian Science Foundation under the project HRZZ-UIP-2017-05-7169.

tinkamų veiksmažodžių išrinkimo ir tikrinimo procesas. Aptariant terminų patikrinimo rezultatus ypatingas dėmesys skiriamas veiksmažodžių ir žodžių junginių vaidmeniui, nustatomi tinkami semantiniai freimai ir jų elementai.

ESMINIAI ŽODŽIAI: aviacijos terminija, anglų ir kroatų kalbų tekstynas, semantiniai freimai, specializuotos žinios, terminologija.

INTRODUCTION

There have been numerous applications of the theory of Frame Semantics (Fillmore 1976, 1982) in lexicography and computational linguistics, most known of which is certainly the lexical knowledge database FrameNet (Fillmore et al. 2003; Ruppenhofer et al. 2017). The work on mapping relations between syntax and semantics, which FrameNet essentially is, has not only inspired similar projects concerning other languages besides English, but has also paved the way for research on specialized knowledge and the organization of specialized knowledge categories based on semantic frames. Regardless of whether they adhere to FrameNet's methodology closely or they apply a modified approach to identifying and delineating semantic frames as types of categories, the projects created around building resources designed for specialized knowledge communities needed to take into account the specifics of terminology and the context of languages for specific purposes as opposed to general language and its use (Faber et al. 2006; 2009; L'Homme 2012).

There has been no lexical resource developed for the Croatian language that is completely based on FrameNet's methodology and semantic inventory. Certain semantic frames have been defined as part of a larger database of conceptual metaphors in Croatian (Despot et al. 2019). However, a database of semantic frames of aviation, which is being developed within the research project the *Dynamicity of Specialized Knowledge Categories* (DIKA, ihjj.hr/dika), is the first attempt at building such a specialized database for Croatian. The project deals with the description of conceptual and linguistic levels of specialized knowledge categories within the dynamic nature of semantic frames. The project's key output is a multilingual terminological database in which aviation terminology is defined in semantic frames, including frame elements, conceptual relations, and figurative and phraseological terminological units. A parallel corpus of English and Croatian texts in the domain of air traffic has been compiled for the purposes of semantic and syntactic analysis, and it will serve as

the starting point for term extraction and term analysis. The subdomain of air traffic has been chosen as the most representative aspect of the vast and interdisciplinary field of aviation.

This paper discusses the idea of using parallel corpora for the extraction of terminological units referring to specific semantic frames of aviation, as the first step in developing a method for determining semantic frames and their conceptual relations in the field of aviation. Although parallel corpora have long been recognized as invaluable sources of term candidates for building various specialized resources (Vintar 2000), they have so far been little used in applied research referring to Frame Semantics.² The process of compiling the parallel corpus of texts taken from the Eur-Lex database using Sketch Engine tools is described first. The methodology of extracting term candidates and their manual verification is then presented, as well as the process of extracting and verifying relevant verbs. The results of term verification are presented and commented on, followed by the discussion of the potential use of verbs and verbal collocations in identifying relevant semantic frames and frame elements in the field of aviation.³

1. THEORETICAL BACKGROUND

In contemporary terminology and LSP communities, specialized knowledge resources are no longer built around the description of hierarchical relations only because it is considered they do not reflect the dynamic nature of the knowledge categories (Faber et al. 2006; 2009; L'Homme, Robichaud 2014). Moreover, the formal distinction between a dictionary and a lexicon that the whole traditional terminography is based on seems to have been blurred with more and more specialized resources employing the methods of compilation and knowledge presentation otherwise used in creating general language resources (L'Homme et al. 2016; L'Homme 2018). Lexical and semantic resources built on the premises of Frame Semantics define knowledge in terms of semantic frames as a type of dynamic categories of intricate elements and mutual relations.

Each semantic frame is defined as a type of event or state with its participants, called frame elements (FEs). Semantic frames are invoked by

² Parallel corpora would certainly reach their full potential in the creation of multilingual lexical repositories based on semantic frames (Boas 2005).

³ We thank the anonymous reviewers for their invaluable comments and suggestions.

lexical units, i.e. words in everyday language or terms in specialized knowledge, which instantiate frame elements. The basic criterion for delimiting one frame from another is that all lexical units should “evoke the same type of event and share the same inventory and configuration of FEs” (Ruppenhofer et al. 2017: 384–385). This method for arriving at the scope of semantic frames and their definitions could be, however, considered an arbitrary one because it does not require a previous “overview” of the available lexical data. Since many frames follow a scenario-like composition, their description usually starts from selecting relevant verbs (Ruppenhofer et al. 2010) to which appropriate arguments are added following the annotation of examples extracted from the corpus. Although lexical units and frame elements are defined according to the analysed corpus data, semantic frames are nevertheless mostly intuitively identified and named. In identifying semantic frames of a certain specialized knowledge domain, it would be more suitable to first delineate the scope of knowledge being defined, which in terminology work is done through the identification of key concepts and conceptual relations in the domain.

Terminological research has closely kept track of developments in linguistics over the past two decades, and has shifted its theoretical focus to analysing processes and events as more dynamic specialized knowledge structures. Therefore, it has become clear that predicative terms (verbs, adjectives, adverbs and certain nouns) often denote key concepts of a domain terminology (L’Homme 1998; Pimentel 2012; L’Homme, Robichaud 2014). For instance, terms *communicate*, *fly*, *take off*, and *transport* are typically present in the field of aviation, as well as terms for a number of procedures that we ontologically define as processes or activities. Although the notion of *termhood* is rather well understood among terminologists when it comes to determining the conceptual level of terms, the criteria for choosing which types of lexical units should be included as terms in a terminological resource still differ depending on the field analysed and the theoretical framework applied. The criteria for defining types of verbs carrying specialized meaning proposed by L’Homme (1998, 2015) are most referred to in terminological literature, along with Lorente’s criteria and her typology of verbs in specialized discourse (2002, 2007).

L’Homme distinguishes among three types of verbs in specialized texts: verbs specific to a certain field (e.g. *land*, *take off*), verbs that acquire specialized meaning in the field (e.g. *operate*), and verbs of general lan-

guage.⁴ Verbs that belong to the first and second types can be labelled as terms, but verbs of the second type present a more interesting group for terminological analysis because there are cases when it is not clear whether the meaning is different enough from the verb's canonical or default general dictionary meaning.

The criteria for deciding whether a verb takes on a new, specialized meaning in a given context should yet be more precisely defined. The criteria for deciding whether a verb can be classified as a term depend, according to L'Homme (1998, 2015), on its arguments and its morphological and semantical relation to lexical units of the same meaning that are defined as terms. In short, if a verb has arguments that are defined as terms, it is likely to be a term itself, as in the sentence *The pilot flies the helicopter*. Also, if the noun *takeoff* is defined as a term, then the verb *take off* should also be defined as one. We have applied these criteria in our analysis of extracted verbs, of which more is said in section 4.

2. METHODS

In order to develop a valid methodology for identifying specialized semantic frames, several steps had to be taken. First, a parallel English-Croatian corpus was created. The second stage consisted of term extraction and validation of term candidates' lists. The third step included extracting a list of verbs and analysing their collocations bearing specialized meaning.

2.1. Corpus compilation

The corpus is compiled from the Directory of legal acts of the European Union, chapter *Transport policy*, subchapter *Air transport* in English and Croatian. Out of 220 documents from the *Air transport* subchapter, 178 legal acts are taken having both (English and Croatian) language versions. The texts are downloaded from the EUR-Lex database,⁵ and entered into the Sketch Engine's corpus compilation module.⁶ The simplest way to create a parallel corpus in Sketch Engine is to upload data

⁴ L'Homme (2015) further classifies general language verbs into support verbs (after Lerat 2002) and discursive verbs (following Lorente and her classification from 2002). See Pimentel (2017) for a recent elaboration of these two approaches to typology of verbs in specialized texts.

⁵ Available at: <https://eur-lex.europa.eu/browse/directories/legislation.html> [accessed 05.09.2019].

⁶ Available at: <https://www.sketchengine.eu/user-guide/user-manual/corpora/setting-up-parallel-corpora/> [accessed 05.09.2019].

in a tabular format such as a spreadsheet (Excel). Spreadsheets must contain language names in the first row and then aligned segments (word, sentences, or paragraphs) side by side.

Every document (legal act) was processed to get one column for each language. In order to achieve this, bullets and numbering, as well as numerical part of tables in documents were removed. Each language in the source file was processed into a separate monolingual corpus and aligned with the corresponding corpus in the other language.

Table 1. Size of the parallel English-Croatian corpus

	ENGLISH	CROATIAN		ENGLISH	CROATIAN
Tokens	1,151,297	1,059,406	Sentences	72,045	75,638
Words	951,156	855,560	Documents	178	178

Apart from extracting most frequent single-word and multi-word terms used as candidates for semantic frames elements, the corpus can also be used for the extraction of terms, definitions and examples of the linguistic context that will be entered into the terminological database *AirFrame*. The corpus will also serve for the analysis of the syntactic level of terms, i.e. their collocations, phraseological relations and predicate-argument constructions.

2.2. Term extraction and term verification

Automatic term extraction was conducted for each language with the option of extracting a list of 1000 single-word and multi-word keywords. The EUR-Lex English 2/2016 corpus was used as a reference corpus for extracting English single-word term candidates, while the English Web 2013 was used as a reference corpus for extracting multi-word term candidates. Similar options were possible for term extraction in Croatian. The EUR-Lex Croatian 2/2016 corpus served as a reference corpus for extracting more relevant single-word term candidates, while this option was not possible for extracting multi-term keywords. Therefore, the Croatian Web hrWaC 2.2. corpus was used instead.

In order to test the reliability of the method, the first 100 term candidates taken from the English list of multi-word terms were manually validated. Having checked their concordances, out of 100 term candidates, 70 were recognized as valid terms by a terminologist with ample experience in

aviation terminology. The list of English keywords (or single-word terms) consisted largely out of aviation abbreviations and a few key aviation terms.

Manual analysis and term verification of both English and Croatian candidate lists of extracted terms was then conducted. The verification process consisted of two phases: 1. removing all abbreviations and obvious general language words and collocations; and 2. analysing the concordances of the remaining terms in order to verify whether they were indeed aviation terms. Abbreviations are commonly used in aviation as terms, thus not including them in a dictionary or glossary of aviation would be understood as a poor presentation of aviation terminology in use. However, since we wanted to arrive at a list of terms or multi-word constructions that could be used in naming the elements of semantic frames in aviation, abbreviations as lexical substitutes of full terms were not treated in the same function here. Apart from this, there was no reason for keeping abbreviations in the Croatian term candidates lists because they were used in the texts as foreign language elements.

2.3. Extraction and analysis of verbs

The lists of English and Croatian verb were extracted automatically using the *Keywords* option in Sketch Engine. Verbs *be*, *have* and *do* have not been taken into account in the English corpus, while *biti* “be”, *imati* “have”, *moći* “can” and *morati* “must” have been omitted from the analysis in the Croatian corpus. The first 100 verbs from each list are then analysed by manual validation of their word sketches. Collocations deemed as multi-word terms or collocations of relative significance for aviation terminology based on their frequency are noted down in an Excel sheet next to the analysed verb.

3. RESULTS AND DISCUSSION

Manual term verification of single-word and multi-word English and Croatian term candidates yielded surprising results concerning precision and recall of the automatic extraction process in Sketch Engine. For instance, after the first phase of term verification, out of 1000 terms, only 400 single-word Croatian terms remained that were not abbreviations or general language words. The final lists of Croatian terms consist of 199 single-word and 455 multi-word Croatian terms. There are somewhat more English multi-word terms, but not as many as one would expect. 664 multi-word English terms remained after the analysis, but only 148

single-word English terms, which makes for a significantly poor result of only 14.8% valid terms from the initial list of multi-word English terms.

As expected, the second phase of term verification presented several issues regarding the termhood of the candidates. If both the noun and the adjective referring to a certain concept appeared on the list, the noun was kept as a more prototypical term realization.⁷ This was more the case with the Croatian terms since in English a noun and an adjective often have the same morphological form, as in the following examples: *kabina*, *kabinski* “cabin”, *navigacija*, *navigacijski* “navigation, navigational”, *helikopter*, *helikopterski* “helicopter”, *identifikacija*, *identifikacijski* “identification”. Similarly, if both an adjective and an adverb appeared in the lists, the adjective form was kept: *instrumentalni*, *instrumentalno* “instrument”, while in the case of the verb/adjective difference, the verb was normally kept on the list: *certificirati*, *certificiran* “certify, certified”. When two nouns referring to the same concept appeared, both were kept if one of them was referring to the action, and the other to the result, as in the example of *identifikacija* and *identificiranje* “identification”. It seems that no such difference exists in the English counterparts.

The English term candidates lists included a lot of information that needed to be filtered, such as the names of countries, personal names (e.g. Juncker), words in other foreign languages, and even some function words (e.g. the pronoun *this*). The algorithm for the extraction of keywords and multi-word terms apparently recognizes foreign elements, i.e. non-English words in the corpus as terms, too. This could have been avoided had the phrases (in all EU languages) that usually appear at the end of each document been erased prior to adding the documents to the corpus, but it was decided that it was easier to ignore them in the candidates’ lists afterwards.

The second phase term verification regarding the English term candidates showed the inability to distinguish two-part from three-part multi-word constructions as the greatest pitfall of the applied algorithm, which is a common error in many statistical term extraction tools. *Bound air* and *simulated engine* are thus parts of larger, correct multi-word terms *bound air cargo* and *simulated engine failure* (both versions appeared on the list), as well as *instrument ground* and *crew compartment* (instead of

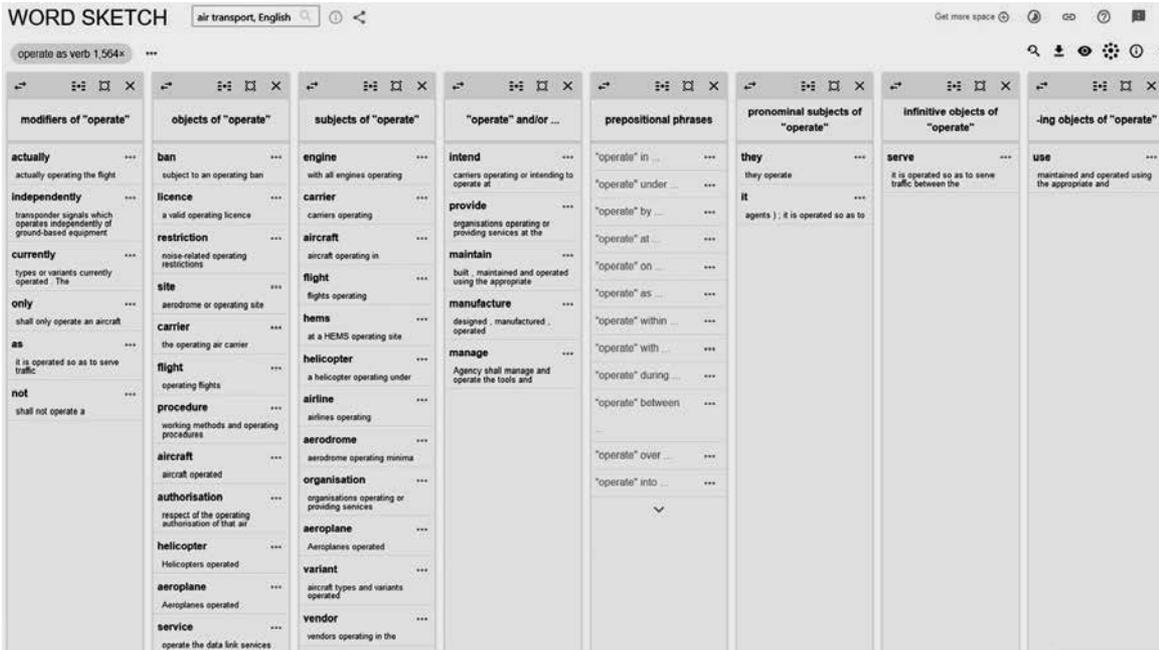
⁷ This kind of reasoning is more in line with traditional terminology work, in which nouns had priority in terminological dictionaries over corresponding adjectives and verbs referring to the same concept (Sager 1990: 58). However, both would be included as lexical units connected with the semantic frame they invoke.

instrument ground time and *flight crew compartment*). Some multi-word term candidates formally appear as proper terms and it takes domain specific knowledge to label them as incorrect or false positive results, e.g.: *visual flight* > *visual flight rules*, *providing air traffic* > *providing air traffic services*, *investigation authority* > *safety investigation authority*, *maximum take-off* > *maximum take-off mass*, *secondary surveillance* > *secondary surveillance radar*. Shorter versions are often simply synonyms, like *controller licence* for *air traffic controller licence*.

Another group of false positive multi-word terms consists of multi-word expressions that could be labelled as belonging to general language: *relevant aircraft*, *individual aircraft*, *specific aircraft*, *large aircraft*, *present aircraft*, *height of cloud base*, *field of civil aviation*, *entire flying time* or *specific training*. Although they are morpho-syntactically valid forms, they refer to no particular concept and therefore cannot be considered terms.

Having analysed all lists, we arrived at a very small number of single-word terms, 199 Croatian and 148 English terms. In order to verify whether the terms that could be considered to denote “key” concepts of the field had been extracted, the word sketches of the first 10 terms in each language were analysed. Word sketches are automatic, corpus-derived summaries of words’ grammatical and collocational patterns, and they serve as an excellent source for generating a large number of term candidates in their syntactical context (Kilgarriff et al. 2014). Apart from typical adjective/noun (*wet-leased aircraft*) or noun/noun (*aircraft noise*, *aircraft position*) collocations, verbal collocations (*operate an aircraft*) and prepositional phrases (*aircraft entering into*) give invaluable terminological information most useful to terminologists and translators. The most relevant relation for this purpose was the coordination relation between the target word and the related words. As can be seen in Figure 1, the words related to aircraft are *occupant*, *terrain*, *vehicle*, *product*, *systems*, *content*, *loss*, *unit*, *engine*, *equipment* and *type*. We compared the list of the coordinated terms with the list of the single-word terms and marked those that appeared in both. The second step in verifying the results was comparing the most frequent collocations of the terms in word sketches with those from the list of multi-word terms. When the collocations from the word sketch (e.g. *aircraft identification* or *aircraft operator* in Figure 1) matched those in the list of terms extracted from the corpus, they were also marked as lexical units to be included in the description of a semantic frame.

Figure 1. Word sketch of the term *aircraft*



The analysis of extracted verbs also consisted of two steps. First, specialized verbs or verbs specific to aviation were manually marked, e.g. *fly*, *land*, *operate*, *communicate*, etc. The second step consisted of marking the verbs that potentially acquire a specialized meaning in the context of aviation. In order to do that, the word sketches of the first 100 most frequent verbs were analysed, from which relevant collocations were further checked for their concordances. Collocations were chosen based on their frequency, but they all had an aviation term as their second element.

When listed by their frequency in the corpus, the largest number of verbs included verbs often referred to as support verbs (Lerat 2002; Pimentel 2012) and phraseological verbs (Lorente 2002; 2007).⁸ Phraseological verbs gain their specialized meaning in syntagmatic units in which at least one terminological unit has the syntactic role of the subject or the object (Pimentel 2012: 87). The verbs like *provide*, *ensure*, *require*, *refer to*, or *establish* are often found in the legislative discourse that makes up the English–Croatian parallel corpus, but whether they bear a special-

⁸ In her redefinition of the classification of verbs appearing in specialized texts (2007), Lorente defines phraseological verbs as those including verbs appearing in collocations and fixed phrases, as well as support verbs.

ized meaning related to aviation can only be judged from their collocations with aviation terms previously extracted. As can be seen in examples (1) and (2), *ensure* and *establish* appear here as parts of arguments, not as predicates in their respective sentences:

- (1) 'Taxiway' means a defined path on a land aerodrome *established* for the taxiing of aircraft and intended to provide a link between one part of the aerodrome and another.
- (2) All documents <...> shall be retained in order to provide the information necessary to *ensure* the continued airworthiness of the aircraft.
- (3) An operating air carrier denying boarding or cancelling a flight *shall provide* each passenger affected with a written notice setting out the rules for compensation and assistance in line with Regulation 261/2004.

Provide, on the other hand, does act as the predicate whose arguments are aviation terms, but does not bear the meaning that could be said to differ from its canonical meaning. While it would be safe to determine collocations *establish aerodrome* and *ensure airworthiness (of the aircraft)* appearing in examples (1) and (2) as collocations with specialized meaning, and verbs *establish* and *ensure* as phraseological verbs in aviation related specialized texts, we cannot classify them as specialized verbs for two reasons. First, they are not the predicates in these sentences, and second, their primary meaning does not change in these syntagmatic units.

Support verbs constitute another subcategory of phraseological verbs that, according to Lorente (2007), cannot be equated with specialized or terminological verbs. Lerat (2002) lists *take* (Fr. *prendre*) as an example of a support verb, one that can appear in collocations with terms, but is not a term itself, as in the collocation *take the flight* (4):

- (4) If no notification is made in accordance with relevant national rules, the managing body shall make all reasonable efforts to provide assistance in such a way that the person concerned is able to *take the flight* for which he or she holds a reservation.

Most extracted verbs, English and Croatian equally, behave in the same manner. However, the phrasal verb *take off* is not only a verb "specific to a certain field" (L'Homme 1998), i.e. specific for the context of aviation. It is one of the very key terms in aviation, equally frequently used as its derivative *takeoff*:

- (5) An aircraft taxiing on the manoeuvring area of an aerodrome shall give way to aircraft *taking off* or about to *take off*.

As for the phrasal verb *give way*, it is not possible to define it as a term solely on the basis of the context or its arguments. According to one of L’Homme’s criteria for determining specialized verbs, a verb is likely to be a term if “the linguistic realizations of actants (i.e. arguments) of a verb are defined as terms” (L’Homme 2015: 80). In example (5), both arguments are in fact terms, but the meaning of the phrasal verb *give way*, however, is the same as in a more general context. Therefore, it wouldn’t be justified to label it as a terminological verb in aviation.⁹

In the first 100 most frequent verbs in the English part of the aviation parallel corpus, only two verbs can be said to be proper specialized verbs or verbs specific for the field of aviation, *operate* and *fly*. The verb *operate* appears in a number of collocations or larger syntagmatic units in which it often appears in other morphosyntactic functions (*operating air commercial flights*, *operate the route*, *carriers operating passenger flights*, *flights operating*), with some of them being recognized as terms (*aerodrome operating minima*). Given the discourse of the corpus texts, verbs like *communicate*, *land*, *taxi* or *transfer* appear in less frequent verbs.

In the first 100 most frequent Croatian verbs, the verb *letjeti* “fly” appears as the only specialized verb. However, given that it is also a verb of general language that keeps the same meaning in a specialized context when used in the predicative function, its arguments and collocations are more relevant in the analysis of predicative aviation terms (e.g. *letovi koji lete* “flights that fly” or *fly the aircraft* in English).

CONCLUSION

The work presented in the paper hopefully points to the conclusion that terminological and linguistic data extracted from specialized parallel corpora can be well used in setting the foundation for defining specialized semantic frames. Although term validation has underlined certain pitfalls of term extraction from corpora compiled of EU legislation, such as the influence of legal discourse on the domain terminology, it has nevertheless proved that parallel corpora remain a reliable source for quickly obtaining term equivalents in other languages.

The analysis of extracted verbs and their classification according to several established typologies of verbs in specialized discourse (L’Homme

⁹ The conclusion stands provided we only take *aircraft* to be the term here, since a whole clause – regardless of the fact it contains several aviation terms – cannot be viewed as a term.

1998; Lerat 2002; Lorente 2007) also yielded interesting, but at the same time somewhat disparaging results concerning the potential termhood of the extracted candidates. Although the number of specialized verbs or verbs specific to the field of aviation was very low on the list of potential candidates, collocations of those verbs showed to be terminologically relevant because they were either multi-word terms or verbal collocations of terms. As the *AirFrame* database lists phraseological terminological units like term collocations, word sketches of given verbs provided many relevant terminological units. The conducted verb analysis points to the conclusion that using verbs as a starting point in defining specialized semantic frames of aviation might be a less fruitful option than analysing word sketches. More relevant terminological data was gained by first determining key entities in the field, and then defining the processes in which they participate by analysing their collocations. Verbs appearing in these collocations should refer to the most relevant processes and events in aviation. Rajh and Grčić Simeunović (2019) made similar conclusions in their paper on classifying the combinatorial behavior of verbs in the marketing domain. Although based on different theoretical approach, their work supports the idea that relevant specialized verbs in a certain domain are best identified through the analysis of the syntactic-semantic behavior of the domain's key concepts, most notably entities.

Having extracted lexical units and validated aviation terms, the next step in identifying aviation semantic frames consists of grouping relevant concepts that the terms refer to, and naming the larger categories connected to them, i.e. frames they invoke. Therefore, lexical units of the same semantic type, and appearing as arguments of the same verbs should be grouped and connected by a frame. Determining the scope of the semantic frame depends largely on the granularity of the conceptual description, as well as on the general idea of the potential use of the planned database. If specialized semantic frames should one day be merged in another lexical resource with frames defining general language, then more abstract, top-level categorization should also be included to the description of the specialized categories of the domain.

REFERENCES

- Boas Hans C. 2005: Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. – *International Journal of Lexicography* 18(4), 445–478.
- Despot K., Tonković M., Brdar M., Essert M., Perak B., Ostroški Anić A., Nahod B., Pandžić I. 2019: MetaNet.HR: Croatian Metaphor Repository. – *Metaphor and Metonymy in the Digital Age*. M. Bolognesi, M. Brdar, K. Despot (eds.), Amsterdam: John Benjamins Publishing Company, 123–146.

- Faber P., León Araúz P., Prieto V., Juan A. 2009: Semantic Relations, Dynamicity, and Terminological Knowledge Bases. – *Current Issues in Language Studies* 1, 1–23.
- Faber P., Montero Martínez S., Castro P., María R., Senso Ruiz J., Prieto V., Juan A., León Araúz P., Márquez Linares C., Vega Expósito M. 2006: Process-oriented Terminology Management in the Domain of Coastal Engineering. – *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12(2), 189–213.
- Kilgarriff A., Baisa V., Bušta J., Jakubčíček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V. 2014: The Sketch Engine: Ten Years On. – *Lexicography* 1(1), 7–36.
- Lerat P. 2002: Qu'est-ce que le verbe spécialisé? Le cas du droit. – *Cahiers de Lexicologie* 80, 201–211.
- L'Homme M.-C. 2018: Maintaining the Balance Between Knowledge and the Lexicon in Terminology: A Methodology Based on Frame Semantics. – *Lexicography. Journal of ASIALEX* 4(1), 3–21. Available at: <https://doi.org/10.1007/s40607-018-0034-1>.
- L'Homme M.-C. 2015: Predicative Lexical Units in Terminology. – *Language Production, Cognition, and the Lexicon*. N. Gala, R. Rapp, G. Bel-Enguix (eds.), Springer International Publishing, 75–93.
- L'Homme M.-C. 2012: Adding Syntactico-Semantic Information to Specialized Dictionaries: An Application of the FrameNet Methodology. – *Lexicographica* 23, 233–252.
- L'Homme M.-C. 1998: Le statut de verbe en langue de spécialité et sa description lexicographique. – *Cahiers de lexicologie* 73(2), 61–84.
- L'Homme M.-C., Robichaud B. 2014: Frames and Terminology: Representing Predicative Terms in the Field of the Environment. – *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon*. M. Zock, R. Rapp, C.-R. Huang (eds.), Dublin, 186–197.
- L'Homme M.-C., Subirats C., Robichaud B. 2016: A Proposal for Combining “general” and Specialized Frames. – *Cognitive Aspects of the Lexicon (Cogalex 2016)*. Coling, Osaka, Japan.
- Lorente M. 2007: Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació. – *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Vol. 2: *De deixebles*. M. Lorente, R. Estopà, J. Freixa, J. Martí, C. Tebé (eds.), Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra, 365–380.
- Lorente M. 2004: Construcciones verbales en el discurso de la genómica. Tipología yerbal y discurso científico. – *Studia Romanica Posnaniensia* 31, 353–359.
- Lorente M. 2002: *Verbos y discurso especializado*. Available at: <http://elies.rediris.es/elies16/Lorente.html> [accessed 20.09.2019].
- Pimentel J. 2012: *Criteria for the Validation of Specialized Verb Equivalents: Application in Bilingual Terminography*. Doctoral dissertation, Montreal: University of Montreal, 329.
- Pimentel J. 2017: Specialized Verbs and Specialized Uses of Verbs in a Comparable Corpus of Judgments Produced in Canada, Portugal and Brazil. – *Multiple Perspectives on Terminological Variation*. P. Drouin, A. Francœur, J. Humbley, A. Picton (eds.), Amsterdam/Philadelphia: John Benjamins, 109–130.
- Rajh I., Grčić Simeunović L. 2019: Classification of the Combinatorial Behavior of Verbs in the Marketing Domain. – *Computation and Corpus-Based Phraseology. Europhras 2019, LNAI 11755*. G. Corpas Pastor, R. Mitkov (eds.), Springer, 360–374.
- Ruppenhofer J., Ellsworth M., Petruck M. RL, Johnson Christopher R., Scheffczyk J. 2010: *FrameNet II: Extended Theory and Practice*, Berkeley, California: International Computer Science Institute.
- Sager J. C. 1990: *A Practical Course in Terminology Processing*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Vintar Š. 2000: Using Parallel Corpora for Translation-oriented Term Extraction. – *Suvremena lingvistika* 49–50(1), 143–152.

LEKSIŅŪ VIENETŪ IŠRINKIMAS SPECIALIZUOTIEMS SEMANTINIAMS FREIMAMS NUSTATYTI

Santrauka

Kokių nors kroatų kalbai pritaikytų specializuotų leksikos šaltinių, grindžiamų semantinių freimų ir *FrameNet* metodų taikymu, nebuvo, todėl įgyvendinant tiriamąjį projektą *Dynamicity of Specialized Knowledge Categories* (DIKA, ihj.hr/dika) (liet. k. *Specializuotų žinių kategorijų dinamiškumas*) bus plėtojama aviacijos semantinių freimų

duomenų bazė. Projekto pagrindinis rezultatas – daugiakalbė terminų duomenų bazė, kurioje aviacijos terminija apibrėžiama semantiniiais freimais, įskaitant freimų elementus, konceptualius ryšius ir terminologinius vienetus. Tam tikslui sudarytas lygiagretusis orlaivių eismo anglų ir kroatų kalbų tekstynas – terminų išrinkimo ir analizės atskaitos taškas.

Straipsnyje svarstoma, kaip lygiagretieji tekstynai gali būti naudojami terminijos vienetais, taikomiems aviacijos semantiniams freimams nustatyti, atrinkti. Tai pirmoji metodo specializuotiems semantiniams freimams nustatyti kūrimo pakopa. Pirmiausia aprašoma, kaip, remiantis *Eur-Lex* duomenų baze ir naudojantis *Sketch Engine* priemonėmis, rengiamas paralelinis tekstų rinkinys. Tada pateikiami terminų variantų išrinkimo ir jų patikrinimo rankiniu būdu metodai bei tinkamų veiksmažodžių išrinkimo ir tikrinimo procesas. Aptariant terminų patikrinimo rezultatus ypatingas dėmesys skiriamas veiksmažodžių ir žodžių junginių vaidmeniui, nustatomi tinkami semantiniai freimai ir jų elementai.

Nors į terminus pretenduojančių aviacijai būdingų veiksmažodžių labai mažai, tokių veiksmažodžių kolokacijos terminologiniu požiūriu aktualios, nes jos yra arba iš kelių žodžių sudaryti terminai, arba iš žodžių junginio sudaryti terminai. Taigi, veiksmažodžių analizė leidžia daryti išvadą, kad aktualius terminologijos duomenis geriausiai kaupti pirmiausia nustatant pagrindinius srities subjektus, o vėliau, analizuojant jų junginius, apibrėžti procesus, kuriuose jie dalyvauja.

Gauta 2019-09-24

Ana Ostroški Anić
Institute of Croatian Language and Linguistics
Ulica Republike Austrije 16
HR-10000 Zagreb
E-mail aostrosk@ihjj.hr

Maja Lončar
Institute of Croatian Language and Linguistics
Ulica Republike Austrije 16
HR-10000 Zagreb
E-mail loncarm@ihjj.hr

Martina Pavić
Institute of Croatian Language and Linguistics
Ulica Republike Austrije 16
HR-10000 Zagreb
E-mail mpavic@ihjj.hr